DOCUMENT RESUME

ED 311 054 TM 013 848

AUTHOR Perkins, Kyle; And Others
TITLE The Determination of Hierarchies among TOEFL Vocabulary and Reading Comprehension Items.
PUB DATE 88
NOTE 19p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS English (Second Language); Factor Analysis; *Foreign Students; Item Analysis; *Language Tests; Latent Trait Theory; Prerequisites; *Reading Comprehension; Reading Tests; Text Structure; *Vocabulary Skills
IDENTIFIERS *Item Hierarchies; Ordering Theory; *Test of English as a Foreign Language

ABSTRACT

A study was undertaken to identify the prerequisite relations (or hierarchies among the items) existing in the item responses of a sample of 86 foreign students who took the Test of English as a Foreign Language (TOEFL) vocabulary and reading comprehension test, Form 3JTF1. The form contains 30 vocabulary items and 30 reading comprehension items. The average TOEFL total score for these students was 463.19, with a standard deviation of 59.96. The average vocabulary and reading comprehension scaled score for these students was 44.60, with a standard deviation of 7.03. Results were submitted to ordering-theoretic analyses. Four significant prerequisite relations were described in terms of their frequency of occurrence and distribution across subject matter categories. Text structure, facet theory, and comprehension category characteristics were used to describe the reading comprehension items found in the significant prerequisite relations. Factor-analytic procedures indicated that there was only one content factor underlying the data to account for the paucity of prerequisite relations in the data set. Three figures are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED311054

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

KYLE PERKINS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Kyle Perkins,[1] Charles Parish, and John T. Pohlmann

Southern Illinois University

Department of Linguistics
Southern Illinois University
Carbondale, IL 62901-4517

[1]The first author gratefully acknowledges the assistance of Jacqueline Ross and Stella R. Cowell of Educational Testing Service who provided the data set for this research.

TM013848

The Determination of Hierarchies Among TOEFL Vocabulary and Reading Comprehension Items

Abstract

Item responses from a TOEFL vocabulary and reading comprehension test were submitted to ordering-theoretic analyses. Four significant prerequisite relationships were discovered. The vocabulary items which were members of the significant prerequisite relations were described in terms of their frequency of occurrence and distribution across subject matter categories. Text structure, facet design, and comprehension category characteristics were used to describe the reading comprehension items found in the significant prerequisite relations. Factor-analytic procedures indicated that there was only one content factor underlying the data which accounts for the paucity of prerequisite relations in the data set.

INTRODUCTION

The purpose of this research was to identify the prerequisite relations (or hierarchies among the items) which exist in the item responses elicited from a sample of ESL students who took the TOEFL vocabulary and reading comprehension test, Form 3JTF1. The impetus for conducting this research was to determine whether mastery of one vocabulary or reading comprehension task is a prerequisite to mastery of two or more vocabulary or reading comprehension tasks, or conversely, whether mastery of two or more vocabulary or reading comprehension tasks is a prerequisite to one higher order vocabulary or reading comprehension task. In the context of this research we employ Gagne's (1977) definition of prerequisite: "a capability of prior learning which is incorporated into new learning. The previously learned entity actually enters into the newly learned capability, becomes and remains a part of the behavior which results from the events of learning" (p. 268).

The results of this research were intended to further inform the corpus of information concerning the growth and development of ESL students' reading comprehension proficiency and processing. It is a well-established tenet that ascertaining what the reader brings to the reading comprehension process is important. Gagne notes that previously learned capabilities must be retrievable from long-term memory; must be readily accessible; may support new learning; and may be incorporated into new learning. Thus, the determination of prerequisite relations in a vocabulary and reading comprehension test may provide some insight into the development of reading skills in ESL students.

Ordering Theory

This research employed ordering theory as the data analysis procedure (Airasian and Bart, 1975; Bart and Krus, 1973; Bart, Frey, and Baxter, 1979). Airasian and Bart define ordering theory as "an approach to fundamental measurement which has as its intent the definition of the best fitting linear or non-linear hierarchical network among a series of tasks....a deterministic measurement model which uses task response patterns to identify both linear and non-linear qualitative, prerequisite relations among tasks or behaviors" (p. 166). Ordering theory is a particularly useful tool for learning research because Gagne has reported that groups of tasks are likely to form branched, hierarchical, non-linear networks rather than hierarchical, linear networks.

Bart and his colleagues consider task A to be a prerequisite to task B to the extent that the (01) response pattern does not occur, where 0 represents a "fail" on task A and 1 represents a "pass" on task B. In other words, the (01) response pattern is disconfirmatory to the extent that a correct response for task A is a prerequisite to a correct response for task B. As Bart and Krus have noted, the (00), (10), and (11) response patterns are confirmatory, while the (01) response pattern is disconfirmatory for the defining statement specifying the prerequisite relationship: success on task A is a prerequisite to success on task B if and only if the response pattern (01) for tasks A and B respectively does not occur.

There are two parameters which the data must meet for an ordering-theoretic analysis. First, all tasks must be scored dichotomously, and second, all subjects in the sample must respond to all tasks. There is the further assumption that a correct response to a task reflects a manifestation of the latent trait being assessed.

METHOD

Subjects

The subjects for this study were 86 foreign students who were enrolled in an intensive English center. The method of selection was the natural assembly process. All students who sat for the examination were included in the sample. The average TOEFL total score for these students was 463.19 with a standard deviation of 59.96. Their average vocabulary and reading comprehension scaled score was 44.60 with a standard deviation of 7.03.

Instrumentation

Form 3JTF1 of the TOEFL vocabulary and reading comprehension test contains 30 vocabulary items and 30 reading comprehension items. There was one experimental item in each section; therefore, only 58 items were entered into the analysis.

Procedures

A contingency table for the 3,306 (58 x 58 - 58 = 3,306) task/item pairs was constructed, and a zero tolerance level was established. In this context, zero tolerance means that if, for any task pair generically labelled AB, a single occurrence of an (01) response pattern was found, the prerequisite relationship that success on task A is necessary for success on task B was disconfirmed for that particular task/item pair.

Once all the prerequisite relations were identified, they were submitted

to a statistical test developed by Bart and Read (1984) to determine whether the prerequisite relationships occurred by chance. Bart and Read noted that the ordering-theoretic analysis is insensitive to item difficulty and that "if item i is an extremely easy item and item j is an extremely difficult item, then intuitively one would expect that most respondents would get item i correct and item j incorrect. This pairing of item difficulties could lead to an inflated occurrence of allowable response patterns (00, 10, and 11) and a deflated occurrence of the disallowable response pattern (01), producing the finding of a prerequisite relationship, when in fact no relationship existed" (p. 224).

Bart and Read's statistical test uses the table of notation for response patterns (Appendix A) and the Fisher's exact test's unit normal deviate approximation, the formula for which is found in Appendix B. The unit normal deviate approximation is in reality a z value, and the criterion level for significance was set in advance at the .05 level. In other words, the test statistic produced by the formula in Appendix B for each prerequisite relationship had to equal or exceed 1.645 in order to be considered significant.

(Put Appendices A and B about here)

Information from the Carroll, Davies, and Richman (1971) word frequency list was used as the framework for describing any vocabulary items that were members of a significant prerequisite relationship found to underlie the item responses. The Carroll et al. frequency list was based on a corpus of 5,088,721 words drawn in 500-word samples from 1,045 texts which reflected a range of required and recommended reading to which students in grades 3 through 9 are exposed. A word's frequency is the number of times the

word type occurred in the corpus of 5,088,721 tokens. The distribution value, having a range of .0 to 1.0, is an index of dispersion based on the dispersion of the frequencies across 17 subject matter categories. The Carroll et al. frequency indices have been utilized successfully in previous ESL research (Perkins and Linnville, 1987).

Meyer's (1975) prose analysis system, Schlesinger and Weiser's (1970) facet design, and Pearson and Johnson's (1978) classification scheme and operational rules for determining different comprehension categories were utilized as the framework for describing the significant prerequisite relations found to underlie the reading comprehension item responses.

Meyer's prose analysis system employs a "content structure" which shows the relationships of all a text's information, and there are three distinct levels of content structure. The top-level structure consists of the rhetorical relationship that binds together all a text's propositions, and it also contains the main ideas which Meyer terms macropropositions. The middle third level of the content structure contains the major supporting details, termed micropropositions, that complement the major premises of the text's message. The bottom third contains very specific, minor details.

Meyer's rhetorical relationships fall into five basic groups:

a. An antecedent/consequent or covariance rhetorical relationship shows a causal relationship between topics.
b. A response rhetorical relationship includes the remark and reply, question and answer, and problem and solution formats.
c. A comparison relationship points out differences and similarities between two or more topics.
d. A collection relationship shows how ideas or events are related together into a group on the basis of some commonality.
e. A description relationship gives more information about a topic by presenting attributes, specifics, explanations, or settings (Meyer and Rice, 1984, pp. 326-327).

Meyer's prose analysis system has been employed in ESL behavioral

analyses (Perkins and Brutten, 1988a).

Schlesinger and Weiser's facet design focuses on the explicit relationship of a reading test's items to the prose passages on which the items were based. "The facets include form of information (explicit or implicit), source of information (the text, formal instruction, and informal experiences), context (appropriate or not appropriate) and the frequency with which information appears in the source" (Schlesinger and Weiser, 1970, p. 566).

Perkins and Brutten (1988b) used the facet design to investigate the effects of background knowledge in ESL reading comprehension tests.

Pearson and Johnson developed a classification scheme to focus on the relationship between reading comprehension questions and their keyed responses. The following rules are provided for determining the comprehension categories:

> A question-answer relation is classified as textually explicit if both question and answer are derivable from the text and if the relation between question and answer are explicitly cued by the language of the text (p. 163). A question-answer relation is classified as textually implicit if both question and answer are derivable from the text but there is no logical or grammatical cue tying the question to the answer and the answer given is plausible in light of the question (p. 163). Scriptally implicit (comprehension) occurs whenever a plausible nontextual response is given to a question derivable from the text (p. 164).

RESULTS

Prequisite relations were found to exist between 9 of the 3,306 item/task pairs, 4 of which were significant at the .05 level. Table 1 presents the significant prerequisite relationships which were found to underlie the item responses.

(Put Table 1 about here)

The defining statement in the Bart-Krus ordering-theory paradigm is A ⟶ B which is to be read as "success on item A is a prerequisite to success on item B." The "A" item was the first item in the vocabulary and reading comprehension test. One assumes that the first item was the A term in the Table 1 tree because it was the easiest item in the test: 80 of the86 respondants answered the item correctly, giving it an item difficulty of .93. Henning (1987) notes that "many test developers prefer to begin each section of their tests with a few very easy items. This is done to overcome psychological inertia on the part of the examinee" (p. 50).

The "A" item was a vocabulary item, a word type which occurred 637 times in the Carroll et al. total corpus of 5,088,721 tokens. With a dispersion index of .7870, it occurred over a broad spectrum of the 17 subject categories examined by Carroll et al.

The "B" items included two vocabulary items and two reading comprehension items. Both "B" vocabulary items had lower frequency indices, 13 and 55, than the "A" item and more restricted dispersion indices, .3756 and .7437. In sum, the examinees were more likely to have encountered the "A" item than they were likely to have encountered the "B" vocabulary items.

The two "B" reading comprehension items had their text passage foci on the Meyer middle text structure level, the microproposition. One of the reading comprehension items involved a textually explicit factual question; the other, a scriptally implicit inference item which entailed that the subjects utilize textually relevant background knowledge.

Meyer's research indicates that microproposition-based reading compre-

hension questions may be more difficult to process, which could explain why the two "B" reading comprehension items were focused on micropropositions:

> First, ideas which are located at the top levels of a structural analysis of prose are recalled and retained better than ideas which are located at the lower levels....Second, different items of information located high in the structure are more likely to be integrated in memory than items located low in the structure.... Third, the type and structure of relationships among ideas in prose dramatically influence recall when they occur at the top levels of the structure; however, when the same relationships occur low in the structure, they have little effect on recall (Meyer and Rice, 1984, p. 329).

We were somewhat perplexed as to why we found only 4 significant prerequisite relationships in 3,306 item/task pairs; therefore, we sought to find an explanation. A data set can fail to yield prerequisite relations for at least two reasons: (1) the scale in unreliable; (2) only one content factor or latent structure underlies the item responses. We can immediately discount the first possibility because the internal consistency reliability estimate for the TOEFL subtests is in the .8-.9 interval. Thus, we elected to investigate the content factor structure of the TOEFL vocabulary and reading comprehension item responses.

A factor analysis was performed in order to examine the latent structure of the TOEFL vocabulary and reading comprehension test. The inter-item correlation matrix was subjected to a principal components analysis. The number of factors present in the test was estimated using a scree test and a parallel analysis (Horn, 1965). The scree test suggested a one-factor structure.

A parallel analysis was performed by generating a random data matrix of the same order as the present TOEFL data matrix (86 rows and 58 columns). The random variables were independent binomial variates whose expected

values were equal to the item difficulties observed on the TOEFL items. The random inter-item correlation matrix was also subjected to a principal components analysis. The eigenvalues from the TOEFL and random correlation matrices were then compared. Only the first two TOEFL eigenvalues were larger than the corresponding random eigenvalues. The second TOEFL eigenvalue was only slightly larger than the second random eigenvalue. An attempt was then made to interpret a two-factor TOEFL vocabulary and reading comprehension structure.

Two TOEFL factors were rotated to the normal varimax criterion. The two varimax factors were found to be difficulty factors; the easy items tended to load on the first factor and the difficult items tended to load on the second factor. It was therefore concluded that only one general content factor was present in the TOEFL vocabulary and reading comprehension test, since the more complex two-factor structure was confounded with item difficulty. The KR-20 reliability of the general factor was .886.

SUMMARY

We attempted to identify and to describe the significant prerequisite relations underlying the item responses from one form of the TOEFL vocabulary and reading comprehension test; the sample's mean TOEFL total score was 463, and the mean vocabulary and reading comprehension subtest score was 44.60. The 4 significant prerequisite relationships which were identified were described in terms of the vocabulary items' frequency of occurrence and dispersion traits. The reading comprehension items were described in a text structure, facet theory, and comprehension category framework. Factor-analytic procedures indicated that only one content factor underlay the data, thereby accounting for the paucity of prerequisite relationships.

REFERENCES

Airasian, P. W., & Bart. W. M. (1975). Validating a priori instructional hierarchies. Journal of Educational Measurement, 12, 163-173.

Bart, W. M., & Krus, D. J. (1973). An ordering-theoretic method to determine hierarchies among items. Educational and Psychological Measurement, 33, 291-300.

Bart, W. M., Frey, S., & Baxter, J. (1979). Generalizability of the ordering among five formal reasoning tasks by an ordering-theoretic method. Child Study Journal, 9, 251-259.

Bart, W. M., & Read, S. A. (1984). A statistical test for prerequisite relations. Educational and Psychological Measurement, 44, 223-227.

Brownlee, K. (1965). Statistical theory and methodology in science and engineering. New York: Wiley.

Carroll, J. B., Davies, P., & Richman, B. (1971). Word frequency book. Boston: Houghton Mifflin.

Gagne, R. M. (1977). The conditions of learning (third edition). New York: Holt, Rinehart and Winston.

Henning, G. H. (1987). A guide to language testing: development, evaluation, research. Cambridge, MA: Newbury House Publishers.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

Meyer, B. J. F. (1975). The organization of prose and its effect on memory. Amsterdam: North-Holland Publishing.

Meyer, B. J. F., & Rice, G. E. (1984). The structure of text. In P. D. Pearson (ed.), Handbook of reading research (pp. 319-351). New York: Longman.

Pearson, P. D., & Johnson, D. D. (1978). Teaching reading comprehension. New York: Holt, Rinehart and Winston.

Perkins, K., & Linnville, S. E. (1987). A construct definition study of a standardized ESL vocabulary test. Language Testing, 4, 125-141.

Perkins, K., & Brutten, S. R. (1988a). A behavioral anchoring analysis of three ESL reading comprehension tests. TESOL Quarterly, 22, 607-622.

Perkins, K., & Brutten, S. R. (1988b). Using a facet design to assess the effects of the form, the source, and the frequency of information necessary to answer ESL reading comprehension questions. Language Learning, 38, 171-185.

Schlesinger, I. M., & Weiser, Z. (1970). A facet design for tests of reading comprehension. Reading Research Quarterly, 5, 566-580.

Appendix A
Table of Notation for Response Patterns for Two Dichotomously Scored Items, A and B

		Item B		
		1	0	
Item A	1	n_{11}	n_{12}	$n_{1.}$
	0	n_{21}	n_{22}	$n_{2.}$
		$n_{.1}$	$n_{.2}$	

(Bart and Read, 1984, p. 225)

Appendix B
Unit Normal Deviate Approximation

$$u = \frac{n_{11} - \frac{n_{1.}n_{.1}}{n}}{\left(\frac{n_{1.}n_{2.}n_{.1}n_{.2}}{n^3}\right)^{1/2}}$$

(Brownlee, 1965, p. 165)

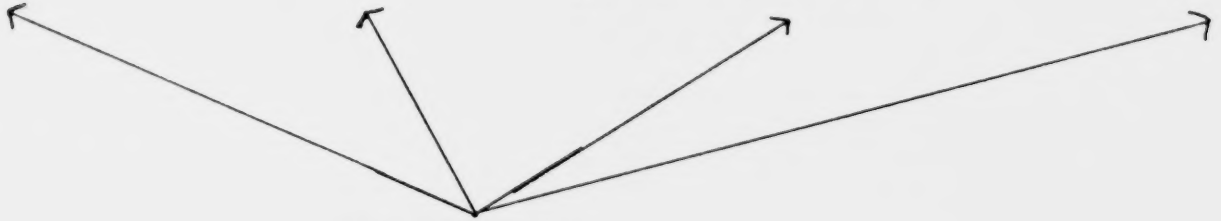
Table 1
Significant Prerequisite Relationships

a vocabulary item: an abstract noun; frequency = 13; dispersion = .3756

a vocabulary item: an abstract noun; frequency = 55; dispersion = .7437

a reading comprehension item: the recognition of a stated factual detail; a description rhetorical organization; a textually explicit microproposition; explicit form of information; textually explicit source of information appeared once in the text

a reading comprehension item: the inference of an underlying, unstated relationship; a description rhetorical organization; a scriptally implicit microproposition; implicit form of information; source of information was nontextual



a vocabulary item: an attributive adjective; frequency = 637; dispersion = .7870